UNIVERSITI TEKNOLOGI MARA
FINAL EXAMINATION

COURSE	:	STATISTICS FOR BUSINESS AND SOCIAL SCIENCES
COURSE CODE	:	STA404
EXAMINATION	:	JUNE 2019
TIME	:	2 HOURS

INSTRUCTIONS TO CANDIDATES

1. This question paper consists of seven (7) questions.
2. Answer ALL questions in the Answer Booklet. Start each answer on a new page.
3. Do not bring any material into the examination room unless permission is given by the invigilator.
4. Please check to make sure that this examination pack consists of :
   i) the Question Paper
   ii) a five – page Appendix 1
   iii) an Answer Booklet – provided by the Faculty
   iv) a Statistical Table – provided by the Faculty
5. Answer ALL questions in English.

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO

This examination paper consists of 6 printed pages

© Hak Cipta Universiti Teknologi MARA

QUESTION 1

A researcher wishes to study students' satisfaction (strongly disagree=1, disagree=2, neutral=3, agree=4, strongly agree=5) towards the services provided by the Academic Affairs in Medical College X. The researcher chooses only 10 out of 50 classes. All the students from these 10 classes will be used for the study.

a) Identify the population for the above study.

(1 mark)

b) State the sampling frame for the above study.

(1 mark)

c) Name the variable of interest for the above study. State its type and its level of measurement.

(3 marks)

d) State the sampling technique used in this study. Explain your answer in the context of the study.

(2 marks)

QUESTION 2

The stem and leaf plot below represents the Mathematics test scores (out of 100) of 15 randomly selected students in Class Angle.

Stem	Leaf
4	3
5	2 4 8
6	2 3 5 7 8 8 9
7	1 5 8
8	9

Key : 4 | 3 means 43

a) Calculate the mean and standard deviation for the test score.

(4 marks)

b) Interpret the mean obtained in a).

(1 mark)

© Hak Cipta Universiti Teknologi MARA

c) The summary statistics of the test score for Mathematics and Biology students in Class Angle are summarized in the following table. Using an appropriate measure, determine which distribution of the test score between the subjects is more dispersed.

Descriptive statistics

	N	Mean	Std. Deviation
Mathematics	15	65.4667	11.1859
Biology	17	66.9412	17.5623

(3 marks)

QUESTION 3

A diet trainer wants to determine whether there is a significant difference in the mean weight loss among four diet programs (low calorie diet, low fat diet, low carbohydrate diet and a control group). The data were collected and the result of the weight loss is shown below.

ANOVA

	Sum of Squares	df	Mean Square	F
Between Groups	**A**	3	22.317	**C**
Within Groups	41.600	**B**	2.600	
Total	108.550	19		

a) Find the values of **A**, **B** and **C**.

(3 marks)

b) State the null and alternative hypotheses for this study.

(2 marks)

c) Using *F-test*, is there sufficient evidence to indicate that there is a difference in mean weight loss among the four diet programs? Use $\alpha = 0.05$.

(3 marks)

QUESTION 4

A manager at Company Brilliant wishes to determine whether the employees' work satisfaction is related to their respective department. The results obtained are shown below.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**D**	2	0.541
Likelihood Ratio	2.522	3	0.471
Linear-by-Linear Association	1.314	1	0.252
N of Valid Cases	93		

© Hak Cipta Universiti Teknologi MARA

Satisfaction * Department Crosstabulation

			Department				Total
			Finance	Sales	Human Resource	IT	
Satisfaction	Satisfied	Count	12	38	5	8	63
		Expected Count	12.9	38.6	5.4	6.1	63.0
	Not Satisfied	Count	7	19	3	1	30
		Expected Count	6.1	**E**	2.6	2.9	30.0
Total		Count	19	57	8	9	93
		Expected Count	19.0	57.0	8.0	9.0	93.0

a) Using an appropriate formula, compute the values of **D** and **E.**

(5 marks)

b) State the null and alternative hypotheses for this study.

(1 mark)

c) Based on the *p-value*, is there sufficient evidence to conclude that employees' work satisfaction is related to their respective department? Use $\alpha = 0.05$.

(3 marks)

QUESTION 5

A statistics lecturer intends to investigate whether there is sufficient evidence to conclude that the average score was different from the expected average score of 74. A random sample of 15 students were selected and analyzed using SPSS. The result obtained is as follow.

	N	Mean	Std. Deviation	Std. Error Mean
Exam Score	15	74.4870	13.5564	3.5002

a) Identify the statistical test used for this study.

(1 mark)

b) Show that the standard error of the mean is 3.5002.

(2 marks)

c) Construct a 95% confidence interval for the average score.

(4 marks)

© Hak Cipta Universiti Teknologi MARA

d) Based on the confidence interval in c), does the average score differ from 74? Give a reason to support your answer.

(2 marks)

QUESTION 6

One indicator of physical fitness is resting pulse rate. Ten men volunteered to test the exercise device as advertised on television by using it three times a week for 20 minutes. Their pulse rate (beats per minute) were measured before and after six week of the test. The data were recorded and analyzed using SPSS. The results are shown below.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before	85.2000	10	6.89283	2.17970
	After	82.0000	10	5.94418	1.87972

			Pair 1
			Before - After
Paired Differences	Mean		**G**
	Std. Deviation		2.14994
	95% Confidence Interval of the Difference	Lower	1.66203
		Upper	4.73797
t			4.707
df			9

a) Find the value of **G**.

(1 mark)

b) Prove the *t-statistic* value for the above study.

(2 marks)

c) State the 95% confidence interval for the mean difference before and after six weeks of the test.

(1 mark)

d) At the 5% level of significance, is there sufficient evidence to indicate that this kind of exercise device could reduce resting pulse rate?

(4 marks)

© Hak Cipta Universiti Teknologi MARA

QUESTION 7

A study is conducted to determine the relationship between the length of advertisement (in seconds) aired on TV and sales (in hundreds of RM) of selected ten companies. The data were recorded and analyzed using SPSS. The results are shown below.

Length of advertisement	16	36	22	6	15	12	5	29	28	21
Sales	48	69	71	14	15	29	20	39	65	30

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.017	10.521		.762	.468
	Length of advertisement	1.683	.494	.769	3.406	.009

a) Using the data given, calculate the correlation coefficient between the length of advertisement and sales. Explain the value obtained.

(4 marks)

b) Based on the output, write the least square regression equation. Hence, interpret the value of the slope.

(2 marks)

c) Calculate the value of the coefficient of determination. Hence, interpret the value.

(3 marks)

d) Predict the sales of the company if the length of advertisement is 10 seconds.

(2 marks)

END OF QUESTION PAPER

© Hak Cipta Universiti Teknologi MARA

SAMPLE MEASUREMENTS

Mean	$\bar{x} = \frac{\sum x}{n}$
Standard deviation	$s = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]}$ or $s = \sqrt{\frac{1}{n-1}\left[\sum (x-\bar{x})^2\right]}$
Coefficient of Variation	$CV = \frac{s}{\bar{x}} \times 100\%$
Pearson's Measure of Skewness	Coefficient of Skewness = $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$ OR $\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$

© Hak Cipta Universiti Teknologi MARA

CONFIDENCE INTERVAL

Parameter and description	A (1 - α) 100% confidence interval
Mean μ, for large samples	$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$
Mean μ, for small samples, variance σ^2 unknown	$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$; df = n – 1
Difference in means of two normal distributions $\mu_1 - \mu_2$, variances $\sigma_1^2 = \sigma_2^2$ and unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$; df = $n_1 + n_2 - 2$ $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
Difference in means of two normal distributions $\mu_1 - \mu_2$, variances $\sigma_1^2 \neq \sigma_2^2$ and unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$; $df = \frac{\left[s_1^2/n_1 + s_2^2/n_2 \right]^2}{\frac{\left(s_1^2/n_1 \right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2 \right)^2}{n_2 - 1}}$
Mean difference of two normal distributions for paired samples, μ_d	$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$; df = n – 1 where n is no. of pairs

© Hak Cipta Universiti Teknologi MARA

HYPOTHESIS TESTING

Null Hypothesis	Test statistic
$H_0 : \mu = \mu_0$ σ^2 known, large samples	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ or $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
$H_0 : \mu = \mu_0$ σ^2 known, small samples	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$; $df = n - 1$
$H_0 : \mu_1 - \mu_2 = 0$ $\sigma_1^2 = \sigma_2^2$ and unknown	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$; $df = n_1 + n_2 - 2$ $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
$H_0 : \mu_1 - \mu_2 = 0$ $\sigma_1^2 \neq \sigma_2^2$ and unknown	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $df = \frac{\left[s_1^2/n_1 + s_2^2/n_2\right]^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$
$H_0 : \mu_d = 0$	$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$; $df = n - 1$, where n is no. of pairs
Hypothesis for categorical data	$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$

© Hak Cipta Universiti Teknologi MARA

ANALYSIS OF VARIANCE FOR A COMPLETELY RANDOMIZED DESIGN

Let:

k = the number of different samples (or treatments)

n_i = the size of sample i

T_i = the sum of the values in sample i

n = the number of values in all samples

= $n_1 + n_2 + n_3 + \ldots$

$\sum x$ = the sum of the values in all samples

= $T_1 + T_2 + T_3 + \ldots$

$\sum x^2$ = the sum of the squares of values in all samples

Degrees of freedom for the numerator = k – 1
Degrees of freedom for the denominator = n – k

Total sum of squares: $SST = \sum x^2 - \frac{(\sum x)^2}{n}$

Between-samples sum of squares:

$$SSB = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \ldots \right) - \frac{(\sum x)^2}{n}$$

Within- samples sum of squares = SST - SSB

Variance between samples: $MSB = \frac{SSB}{(k-1)}$

Variance within samples: $MSW = \frac{SSW}{(n-k)}$

Test statistic for a one-way ANOVA test: $F = \frac{MSB}{MSW}$

© Hak Cipta Universiti Teknologi MARA

SIMPLE LINEAR REGRESSION

Sum of squares of *xy*, *xx*, and *yy*:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad \text{and} \quad SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

Least Square Regression Line:

$Y = a + bx$

Least Squares Estimates of *a* and *b*:

$$b = \frac{SS_{xy}}{SS_{xx}} \quad \text{and} \quad a = \bar{y} - b\bar{x}$$

Total sum of squares: $SST = \sum y^2 - \frac{(\sum y)^2}{n}$

Linear correlation coefficient: $r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$

© Hak Cipta Universiti Teknologi MARA